

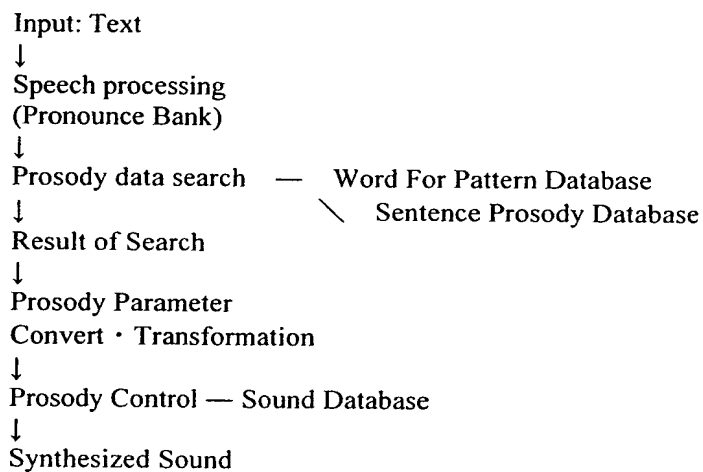
2
B3

"On the control of prosody using word and sentence prosody database" Nukaga N., et al., 1998 Meeting of ASJ Society of Japan, 227-8, 1998

Abstract: This Prosody Control system searches the data whose prosody parallels input-text data from the prosody database. And this system makes synthesized sound using the prosody parameter calculation system.

The flow-chart of the system is described below.

[The prosody control system using word and sentence prosody database]



1-7-24

単語および文韻律データベースを用いた韻律制御方式の検討*

◎額賀信尾

安藤ハル

矢島俊一†

北原義典

((株)日立製作所中央研究所

†(株)日立製作所半導体事業部)

1. はじめに

規則合成音の自然性向上のため、肉声から抽出した韻律パラメータを利用する韻律制御方式の有効性が確認されている。例えば、定型文に関する F_0 ボタンを藤崎モデルのパラメータとして記憶し生成を行う方法[1]や、付属語列の韻律データを用いて韻律を制御する方法[2]などが提案されている。これらの方式では、肉声から抽出した韻律パラメータを利用するため、自然性の高い合成音を生成できる可能性があるが、主として F_0 ボタンの制御を目的としており、時間構造情報についてはさらに検討する必要がある。

本稿では、肉声から抽出した単語及び文に関する F_0 ボタンと継続時間長の韻律パラメータを蓄えたデータベース(以下、韻律DB)から、入力テキストに最も近い韻律を持つデータを検索し、検索されたデータの韻律パラメータを加工することで、入力テキストに対する韻律パラメータを計算する韻律制御方式を提案する。

まず、2章では、本制御方式の基本的な考え方の妥当性を検証するための予備実験を行う。3章で、韻律DBの構成と、 F_0 ボタン及び継続時間長パラメータの計算方法を説明し、本方式の有効性を確認するために行った評価実験について述べる。

2. 予備実験

本章では、文音声から抽出した韻律パラメータを用いた規則合成音が、従来モデルと比較して肉声にどれだけ近いかを確認することにより、本制御方式の基本的な考え方の妥当性を検証する。

実験は、(1)文音声の韻律パラメータの抽出、(2)規則合成音の聴取実験の2段階で行った。

韻律パラメータは、以下の手順で抽出した。

- 文発声音声に対して、目視で音節のセグメンテーションを行う。
- 各音節の継続長、ポーズ長を計測する。
- 各音節の始末端における F_0 をサンプリングし、これらを直線でつないで F_0 ボタンとする。

以上のようにして抽出した韻律パラメータを利用して、以下の3種の女声規則合成音を作成した。合成方式はPSOLA[3]をベースとした。

- X1: 規則で生成した F_0 ボタンと肉声から抽出した継続時間長を付与した合成音
- X2: 規則で生成した継続時間長と肉声から抽出した F_0 ボタンを付与した合成音
- X3: 肉声から抽出した F_0 ボタンと継続時間長を付与した合成音

3種の合成音を用いて、ABX法により聴取実験を行った。すなわち、Xとして上記の3種類の合

成音をランダムに設定し、被験者に聴取させ、肉声(A)及び F_0 ボタン・継続時間長ともに規則により生成した合成音(B)のどちらに近いかを、聴取後に判定させた。文は、交通情報関連の文6文(平均27モーラ/文)である。肉声と合成音は、同一の女性話者の音声を使用した。音源は任意文合成用の素片(VCV)セットである。被験者は、男性8名、女性6名の計14名とした。

実験結果を図1に示す。図中の数字は、各々の判定者数の全体の人数に対する割合を示す。同図より、肉声の F_0 ボタン・継続時間長の両方を付与した合成音は、85%の割合で肉声に近いと判定され、各々単独で付与した合成音よりも、肉声に近いと判定されたことがわかる。

「肉声に近い」と判定 「規則で生成した合成音に近い」と判定

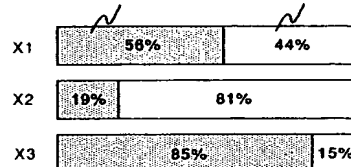


図1 予備実験結果

以上の結果より、(1)文音声から抽出した韻律パラメータを用いると大きく自然性が向上する、(2)継続時間長若しくは F_0 ボタンを単独で用いる場合よりも、両者を利用した方が自然性向上に寄与することが確認できた。

3. 韻律制御方式

3.1. 音声合成システムの概要

図2に、本方式を利用した音声合成システムの処理の流れを示す。合成方式は、2章と同様PSOLAベースの方式を採用している。

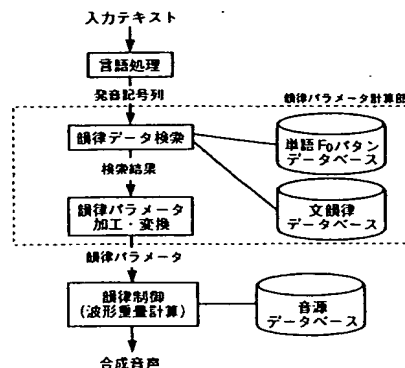


図2 音声合成システムの処理の流れ

* On the Control of Prosody Using Word and Sentence Prosody Database. By N.Nukaga, H.Ando, S.Yajima†, Y.Kitahara (Hitachi Central Research Lab. †Semiconductor Device Center, Hitachi Ltd.)

3.2. 韻律 DB の構成

韻律 DB は、単語 F_0 バタン DB と文韻律 DB から構成される。

3.2.1 単語 F_0 バタン DB

単語 F_0 バタン DB は、モーラ数とアクセント型の組合せに対して 1 つの F_0 バタンを保持しておく。これは、モーラ数及びアクセント型が等しい単語は、 F_0 バタンが類似していることによる。 F_0 バタンは、音節始末端の F_0 による折れ線近似バタンとして表現したものを使用する。

3.2.2 文韻律 DB

文韻律 DB は、文に関する F_0 バタンと継続時間データを保持するものである。現在、数十種の定型文に関する DB を構築している。例えば、ただいま、[地名]から[地名]まで、[状態句]います。

のような文である。便宜上、このような単文を「定型文」と呼ぶことにする。定型文は、可変部と固定部から構成される。上記の例では、[地名][状態句]の部分可変部で、それ以外が固定部である。

例えば、この定型文に合致する入力テキストは、「ただいま、上野から御徒町まで、渋滞しています。」などである。

定型文の F_0 バタンに関する情報としては、以下の 2 つの情報を保持しておく。

- 可変部のアクセント型に対応する固定部の F_0 バタン
- 可変部のアクセント型に対応する可変部の始末端 F_0 データ

また、定型文の継続時間に関する情報としては、以下の 3 つの情報を保持しておく。

- 固定部の音節(音素)継続時間長
- 句間のポーズ長
- 可変部の総時間長

3.3. 韻律パラメータ計算

本節では、単語 F_0 バタン DB と文韻律 DB を利用した韻律パラメータ計算の方法を説明する。

3.3.1 F_0 バタン計算

固定部の F_0 バタンは、可変部のアクセント型に対応した F_0 データをそのまま用いる。

可変部の F_0 バタンは、以下のように求める。

まず、可変部にはいる単語のモーラ数とアクセント型から、単語 F_0 バタン DB を検索して F_0 バタンデータを得る。このままでは、固定部の F_0 との間にギャップが生じるので、可変部の F_0 バタンデ

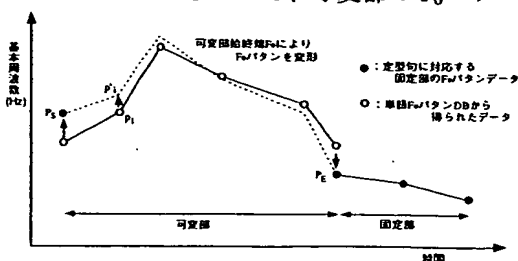


図 3 F_0 バタンの変換例

ータを加工する。本研究では、下記のような変形のための式を用いて F_0 バタンを計算する。

$$p'_i = p_i * \left(\left(1 - \frac{i}{N} \right) * \frac{P_0}{P_S} + \frac{i}{N} * \frac{P_N}{P_E} \right)$$

ここで、 N は可変部にはいる単語のモーラ数、 $p'_i (i=0, 1, \dots, N)$ は求める F_0 データ、 p_i は単語 F_0 バタン DB から得られた F_0 データ、 P_S と P_E は各々、可変部の始端 F_0 データ、終端 F_0 データである。

3.3.2 継続時間長計算

継続時間長の計算は、定型句ごとに行う。定型句は、可変部と固定部からなる句である。

定型句の継続時間長 $D_i (i=0, \dots, N-1)$ は、以下のように求める。ここで、 N は定型句のモーラ数であり、可変部のモーラ数 N_v 、固定部のモーラ数 N_f の和である。

まず、定型句に対応する文韻律 DB 中の継続時間長データを検索し、ポーズ長 DP 、固定部の継続時間長 $D_k (k=0, \dots, N_f-1)$ を得る。続いて、可変部の継続時間長 $D_j (j=0, \dots, N_v-1)$ を、単語継続時間長計算規則を用いて計算する。また、可変部の全継続時間長 DL を、定型句の継続時間長データから得る。次に、この D_j と DL から、次式により、最終的な可変部の継続時間長データ D'_j を計算する。

$$D'_j = D_j * \frac{DL}{\sum_{j=0}^{N_v-1} D_j}$$

3.4. 評価実験

当社従来法との比較評価実験を行った。実験は、従来法及び本方式による韻律を各々付与した女声合成音声をランダムに被験者に聴取させ、抑揚の自然性に関して評価させた。文は、定型文の可変部に地名を埋めこんだ 3 文である。被験者は、男性 8 名、女性 6 名の計 14 名とした。結果は、5 段階評価平均で、従来法 2.7、本方式 3.6 であった。このように、本韻律制御方式により、自然性が大きく向上することが確認された。

4. まとめ

単語及び文韻律データベースを利用した韻律制御方式の検討を行い、合成音の自然性向上に対する有効性を確認した。今後は、効率的な韻律データ収集方法の検討、基本周波数バタン近似方法の検討、任意文への適用を行う予定である。

謝辞 本研究推進にあたって、データ作成及び議論頂いた(株)日立マイコンシステム北爪氏、西川氏、森氏、白戸氏に感謝致します。

参考文献

- [1] 片江他 2 名：文型一韻律データベースを用いた定型文音声合成システム、音講論、pp.275-276、平成 8 年 3 月
- [2] 斎藤他 2 名：付属語連鎖単位を音声合成単位とする規則音声合成、音講論、pp.317-318、平成 6 年 10 月
- [3] F.J. Charpentier and M.G. Stella : Diphone Synthesis using an overlap-add technique for speech waveforms concatenation, ICASSP'86, pp.2015-2018